Journal of Computer Science
https://dergipark.org.tr/en/pub/bbd
Anatolian Science

ISSN,e-ISSN: 2548-1304
Volume:8, Issue:1, pp:24-35 , 2023
https://doi.org/10.53070/bbd.1268221

Research Paper

LBP Feature Extraction and Statistical Pooling-Based Image Spam Detection Model

Aytaç Kaşoğlu[1], Orhan Yaman*[1]

[1]Department of Digital Forensics Engineering, Fırat University, Elazığ, Türkiye

(kasogluaytac44@gmail.com, orhanyaman@firat.edu.tr)

Received:Mar.21,2023 Accepted:May 05,2023 Published:Jun.08,203

Abstract— Email, which stands for electronic mail, is a form of digital communication between two or more individuals. These technological instruments that facilitate communication can have a positive and negative impact on our lives due to junk e-mails, widely known as spam mail. These spam messages, which are typically delivered for commercial purposes by organizations/individuals for indirect or direct benefits, not only distract people but also consume a significant amount of system resources such as processing power, memory, and network bandwidth. In this study, a method based on LBP (Local Binary Patterns) feature extraction and statistical pooling is proposed to classify spam or raw (non-spam) images. Two datasets are used to test the proposed method. The ISH dataset is widely used in the literature and contains 1738 images. In addition to this dataset, the dataset we collect consists of 1015 images in total. Feature extraction was performed on these images. Obtained features were classified by SVM (Support Vector Machine) algorithm. In the proposed method, 98.56% and 79.01% accuracy were calculated for the ISH dataset and our collected dataset, respectively. The results obtained were compared with the studies in the literature.

Keywords : *Spam image detection, Machine learning, LBP feature extraction, SVM.*

1. Introduction

1.1. Background

In the era of information technology, information sharing has enhanced practically every aspect of our lives. Electronic mails are the simplest, fastest, and most rapid method of information sharing over the world. However, due to the convenience of email services, the problem of spam messages has grown increasingly significant, affecting not only people's lives but also consuming a large number of the system's resources. Moreover, these spam emails may contain malicious content, leading to host system security breaches. According to one research, 85 percent of business emails has classified as spam. (Budanović, 2021)

Thanks to developing technologies, people use their e-mails widely. However, spam emails cause harmful content to enter their inboxes, and therefore, a lot of work has been done for spam detection. Spam emails are often text-based, but in some cases, spam can be sent with images, video, and audio files (Kihal & Hamza, 2023). This can fill their inboxes and cause users to be unable to use e-mail services. There are many text-based spam approaches in the literature. But image-based methods are scarce. In this study, a lightweight approach using machine learning is proposed for spam image classification.

1.2. Literature Review

Spam identification has become a demanding research area since spammers are exploring ways to manipulate the information linked to spam phrases by adding complexities. To address such issues, the research employed a variety of machine learning classifiers. This section will discuss the literature review for this research.

Bhuiyan et al. (Bhuiyan vd., 2018), provide an overview of existing email spam filtering methods. They analyze several procedures to summarise diverse spam filtering algorithms and the accuracy of various parameters of different suggested systems. They argue that all present approaches for screening spam emails are effective. Some have had positive outcomes, while others are experimenting with new methods to improve their accuracy performance. Despite their popularity, they all have concerns with spam filtering technologies, which is the key

worry for researchers. (They are attempting to create a next-generation spam detection mechanism capable of comprehending massive amounts of multimedia data and filtering spam emails. They find that the majority of email spam filtering is done using Naive Bayes and the SVM algorithms.

Amara Dinesh Kumar et al. (Kumar vd., 2018), propose a deep learning-based solution for detecting picture spam that employs convolutional neural networks for classification. They have used the convolutional neural network (CNN) which is a deep learning network architecture for image spam detection. The evaluation results show that deep learning surpasses machine learning and other CNN-based methods in terms of accuracy. Çayir, A. et al. (Çayır vd., 2018), Their method combines the convolutional neural network extracting features layers with standard machine learning algorithms such as support vector machines, gradient boosting machines, and random forests. The results reveal that the proposed hybrid models outperform traditional models when trained from raw pixel data. Annadatha, A. et al. (Annadatha & Stamp, 2018), compared two methods for detecting spam images. Initially, they begin by using Principal Component Analysis (PCA) to determine eigenvectors for a set of spam images and then compute scores by projecting images onto the resulting eigenspace. Furthermore, their second approach utilizes Support Vector Machines to extract a broad set of image features and select an optimal subset (SVM).

Singh A.B. et al. (Singh vd., 2022) developed a CNN-based method for analyzing spam images. In the proposed method, they optimized both feature extraction and classification parameters. The developed CNN model was tested with four different datasets, and the results were calculated. They achieved 99.77% success for the ISH dataset. Belkhouche (Belkhouche, 2022), has developed a hybrid model based on CNN that uses both text and images. Byte histogram and 1D-CNN model were applied to five different datasets used in the literature. Accuracy, precision, recall, and F1-measure values for the ISH dataset were calculated as 94.85%, 95.23%, 94.75%, and 94.83%, respectively. Metlapalli et al. (Metlapalli vd., 2022) have classified spam images using the Convolution Neural Network. The performance results of the proposed method were obtained using the Dredze dataset and the ISH dataset. Kihal et al. (Kihal & Hamza, 2023), proposed a CNN-based hybrid model for voice, image, and text spam mail classification. Feature extraction was done with CNN using audio files, video-image files, and text files. The obtained features were combined and classified using the Random Forest algorithm. Accuracy, precision, recall, and F1-measure results were calculated as 98.33%, 98.93%, 97.13%, and 98.02%, respectively, with the recommended method for the images in the ISH dataset. Ghizlane et al. (Ghizlane vd., 2022), proposed a convolutional block attention module (CBAM) model for spam detection in images. They calculated 88.24% accuracy for the ISH dataset with CNN. With CBAM, they achieved 98.65% accuracy for the ISH dataset.

Table 1. Studies and methods used in spam image classification and detection in the literature

Studies	Dataset	Number of Images	Method	Results
(Kumar vd., 2018)	ISH dataset	1738	CNN	Acc = 91%, Rec = 85.7%, Pre = 100%, F1 = 92.3%
(Mahdi Salih & Nadeem Dhannoon, 2020)	ISH dataset	1738	RGB + CNN	Acc = 97.4 %
			HSV + CNN	Acc = 97.4 %
			YCbCr + CNN	Acc = 98 %
			XYZ + CNN	Acc = 98.4 %
			LAB + CNN	Acc = 85.3 %
			YUV + CNN	Acc = 97.8 %
(Annadatha & Stamp, 2018)	ISH dataset	1738	PCA + SVM	Acc = 97 %
(Trivedi, 2016)	Email dataset	-	Bayesian classifier	F1 = 92 %
			NaiveBayes	F1 = 92.8 %

			SVM	F1 = 93.3%
			J48	F1 = 92.1%
			BayesNet (Boosted)	F1 = 92.1%
			NaiveBayes (Boosted)	F1 = 93.2%
(Gao vd., 2008)	ISH dataset	1738	PBT + SVM	Acc = 89.44 % FP = 0.86 %
			SVM	Acc = 80 % FP = 0.86 %
(Rusland vd., 2017)	Spam base and spam data	9324	Modified Naive Bayes	Acc = 83 %
(Singh vd., 2022)	ISH dataset	1738	CNN	Acc = 99.77 %
(Belkhouche, 2022)	ISH dataset	1738	Byte histogram + CNN	Acc = 94.85 %
(Kihal & Hamza, 2023)	ISH dataset	1738	VTA-CNN-RF	Acc = 98.33 %
(Ghizlane vd., 2022)	ISH dataset	1738	CNN with CBAM	Acc = 98.65 %
			CNN without CBAM	Acc = 88.24 %

* Acc=Accuracy, F1=F1 Score, Rec=Recall, Pre=Precision, CNN=Convolutional Neural Network, SVM= Support Vector Machine, VTA = Visual, textual, and audio dataset, RF = Random Forest, CBAM = Convolutional Block Attention Module

As can be seen in Table 1, the ISH dataset is a common dataset used for image spam detection in the literature. CNN-based methods are generally seen in the literature for image spam detection. Deep learning-based models often have high computational complexity. For this reason, (Annadatha & Stamp, 2018), (Trivedi, 2016), (Gao vd., 2008), and (Rusland vd., 2017) developed machine learning-based methods in their studies. In our study, a machine learning-based lightweight method is proposed.

1.3. Motivation and Our Method

Today, the majority of spam emails are texts. For this reason, text-based spam detection methods are quite numerous. But there are not many studies on spam detection from images. Classification of spam images is the main motivation for our study. In the studies in the literature, deep learning-based models or machine learning are generally used for the detection of spam images. Deep learning-based models have high computational power. Computers with high computing power are needed for their implementation. In machine learning-based methods, the accuracy of the method is low as classifiers are used directly. Our motivation is to propose a method that has both low computational complexity and high accuracy. For this reason, feature extraction was performed on the image using the LBP method. The properties obtained with LBP were used together with statistical methods. The obtained results were classified with SVM and high accuracy was calculated. The other motivation of this study is to collect a new dataset besides the ISH dataset, which is widely used in the literature.

Spam detection is a critical issue in computer science and information technology (Zhang vd., 2022). Spam emails not only clog our inboxes, but they may also contain phishing attacks or malicious links, compromising our security and privacy. The volume of spam emails has expanded significantly in recent years, making it difficult for individuals and businesses to efficiently separate spam emails from legitimate ones. One approach to detecting spam emails is through the use of feature extraction and statistical pooling techniques. The Local Binary Pattern (LBP) feature extraction method is a powerful technique for extracting features from images and is effective in a variety of image classification tasks. By combining LBP with statistical pooling techniques, it is possible to effectively extract meaningful features from emails and use them to classify emails as spam or non-spam. There are several reasons why a research proposal on spam detection using LBP feature extraction and statistical pooling would be valuable. First, the development of effective spam detection techniques is important for ensuring the security and privacy of individuals and organizations. By being able to accurately identify and filter out spam

emails, individuals and organizations can protect themselves from phishing attacks and other types of malicious activity. Second, using LBP feature extraction and statistical pooling approaches has the potential to increase the performance of spam detection systems dramatically. These methods have been demonstrated to be effective in a range of picture classification tasks and may be equally beneficial in the context of spam identification. It may be feasible to dramatically enhance the accuracy and efficiency of spam detection algorithms by using these approaches, making them more practical for application in real-world contexts.

1.4. Contributions and Novelties

Contributions and novelties of the proposed method are:

- In this study, a contribution was made to the literature by collecting images for creating a dataset.
- LBP and Statistical based method has been developed for spam detection in images.
- In addition, using the SVM classification algorithm, the spam image detection has been performed more successfully.

2. Materials and Methods

2.1. Materials

Two separate datasets, the "ISH dataset" and "Our collected dataset," were utilized in this proposed study. Figure 1 shows sample photos and the number of images in each dataset.

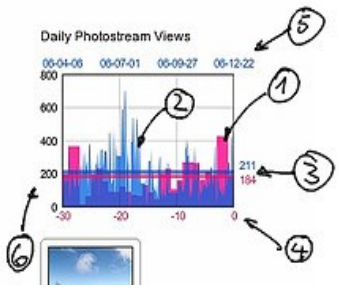
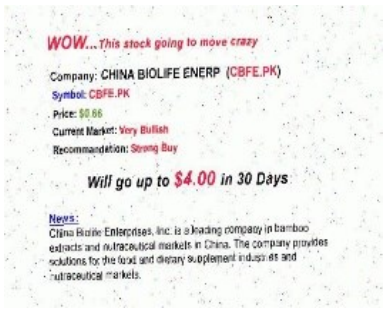
Dataset		Normal image	Spam image
ISH dataset (Abuzaid & Abuhammad, 2022; Ghizlane vd., 2022)	**Sample images**		
	Number of images	810	928
Our collected dataset	**Sample images**		
	Number of images	519	496

Figure 1. Sample images and number of images of datasets

In total, our self-collected dataset contains 1015 images (496 spam images and 519 normal images). These images have been collected at random from "Flickr.com" and "Unsplash.com".

2.2. The Proposed Spam Image Detection Method

The block diagram of the proposed method for the classification of spam images in this study is shown in Figure 2.

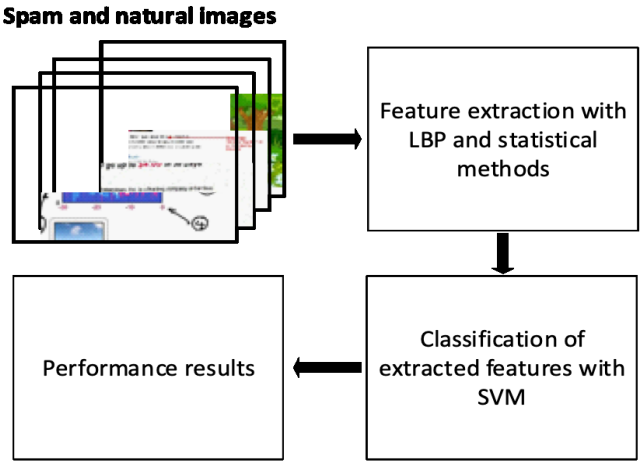


Figure 2. The block diagram of the proposed methods

As shown in Figure 2, the collected images are sent to those dedicated to directly obtaining image features utilizing the Local Binary Pattern (LBP) features extraction method. These features are fed into the classifier during the training step to determine if the image being evaluated is spam or non-spam. This method, which is described in this study, is explained in further detail below.

2.2.1. Local Binary Patterns

The LBP operator was created with the premise that two-dimensional surface textures may be defined by two complementary measures: local spatial patterns and grayscale contrast. The original LBP operator generates labels for image pixels by thresholding each pixel's 3x3 neighborhood with the center value and interpreting the result as a binary integer. (Ojala vd., 2002; PietikÃ¤inen, 2010).

Subsequently, as the approach gained popularity, operations were performed for radius and operators with varying attributes. Equation (1) shows a living representation of the LBP process. In Equation (1), p represents the number of pixels adjacent to the central pixel, x_i $i.$ represents the value of the nearby pixel, and $x_m$represents the value of the center pixel.

$$LBP = \sum_{i=0}^{p-1} s(x_i - x_m)2^i \qquad (1)$$

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

As shown in Equation (1), the threshold is set by comparing the central pixel to each of its neighbors one by one. In threshold measurement, the surrounding pixel is assigned a value of 1 if its value is more than or equal to that of the center pixel, and 0 if it is less.

The threshold measurement reveals an 8-bit LBP code with a total of 256 variations. This code is then mathematically transformed into a numeric number that determines the new value of the center pixel. The weights in the calculation of the specified numerical value are the exponential multiples of two calculated according to the order of the nearby pixels. (Yasar & Ceylan, 2021)

2.2.2. The Statistical Pooling Methods

In this study, additional features were computed utilizing statistical approaches from LBP features. As Statistical methods, “Skewness, Kurtosis, Variance (var), Minimum elements of an array (min), Mean or median

absolute deviation (mad), and Median value of array" functions were used. The mathematical expressions used for Statistical Pooling in this study are given in equation 2-7.

Skewness
$$f1 = \frac{N-1}{(N-2)(N-3)}\left[(N+1)\left(\frac{\frac{1}{N}\sum_{x=1}^{N}(X_i-\overline{X})^4}{\frac{1}{N}\sum_{x=1}^{N}(X_i-\overline{X})^2}-3\right)+6\right] \quad (2)$$

Kurtosis
$$f2 = \frac{\sqrt{N(N-1)}}{N-2}\left(\frac{\frac{1}{N}\sum_{x=1}^{N}(X_i-\overline{X})^3}{\frac{1}{N}\sum_{x=1}^{N}(X_i-\overline{X})^2}\right)^{3/2} \quad (3)$$

Variance
$$f3 = \frac{\sum_{i=1}^{N}(X_i-\overline{X})^2}{N-1} \quad (4)$$

Min
$$f4 = \min\{X_1, X_2, X_3, X_4, ..., X_N\} \quad (5)$$

Mad
$$f5 = \frac{\sum_{i=1}^{N}|X_i-\overline{X}|}{N} \quad (6)$$

Median
$$f6 = \begin{cases} X\left(\frac{N+1}{2}\right) & N \to odd\_number \\ \frac{\left(X\left(\frac{N}{2}\right)+X\left(\frac{N}{2}\right)+1\right)}{2} & N \to even\_number \end{cases} \quad (7)$$

The N value used in Equation 2-7 is the size of the array, and the $\overline{X}$ value is the mean of the values in the array. Following the extraction of image features, the SVM algorithm, one of the machine learning algorithms, classifies them as spam or non-spam.

2.2.3. Support Vector Machines

The Support Vector Machine (SVM) is a well-known Machine Learning classifier. It creates a hyperplane by connecting the closest data points. That marginalizes the classes and increases the distances between them, making it simpler to distinguish between them.

The following are the benefits of support vector machines:

- Effective in high-dimensional environments.
- When the number of dimensions exceeds the number of samples, the method remains effective.
- It also saves memory by using a subset of training points in the decision function (called support vectors).
- The decision function can be provided with several Kernel functions. Common kernels are given, however, custom kernels can also be specified.

SVM classification algorithm was used to classify the statistical features. Classification results were obtained using MATLAB Classification Learner Toolbox. While classifying, 10-fold cross-validation was used for training and testing. The parameters of the Fine Gaussian SVM algorithm used in the proposed method are given in Table 2.

Table 2. Parameters of Fine Gaussian SVM algorithm used in the proposed method

Parameter	Value
Kernel function	Gaussian
Box constraint level	1
Kernel scale mode	Manuel
Manual kernel scale	0.61
Multiclass method	One-vs-one
Standardize	True

Features were extracted with the proposed LBP and statistical pooling-based method. Obtained features were applied with Decision Tree, Linear Discriminant, Naive Bayes, KNN Fine, SVM, AdaBoost, Gradient Boosting, XGBoost, and LightGBM classification algorithms. The accuracy results computed with these machine learning-based methods are shown in Figure 3.

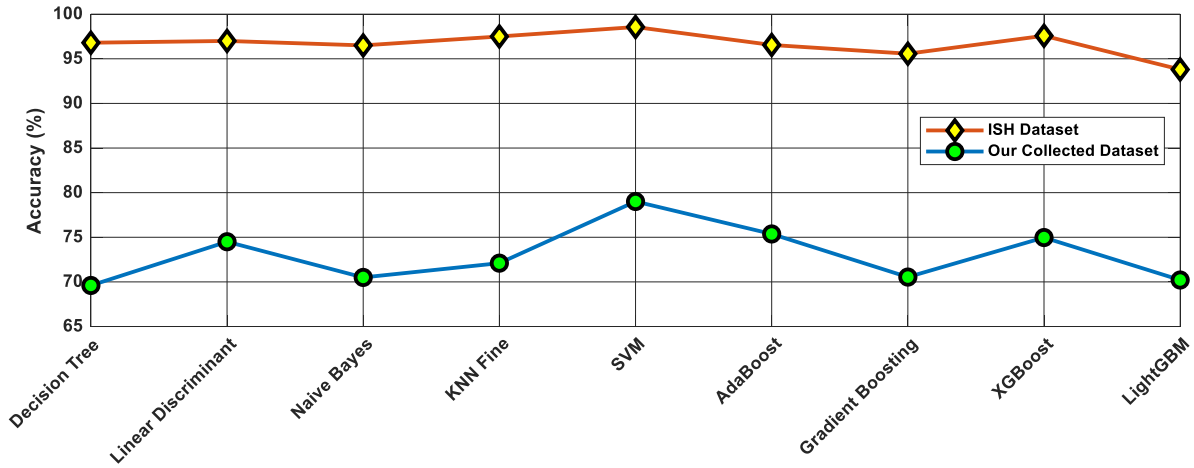


Figure 3. Classification of features obtained by LBP and statistical pooling with machine learning algorithms

As can be seen in Figure 3, the highest classification accuracy was calculated with the SVM algorithm. Therefore, the SVM classifier is proposed in the proposed method.

3. Experimental Results

In this study, four performance measures were used: accuracy, precision, recall, and F-score. The number of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) were used to calculate these performance metrics (FN). Equations show the mathematical representations of the performance measures used. Mathematical notations of the used performance metrics were shown in Eqs. 8-12.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$\Pr ecision = \frac{TP}{TP + FP} \tag{9}$$

$$\operatorname{Re} call = \frac{TP}{TP + FN} \tag{10}$$

$$Geometric\_mean = \sqrt{\frac{TP * TN}{(TP + FN) * (TN + FP)}} \tag{11}$$

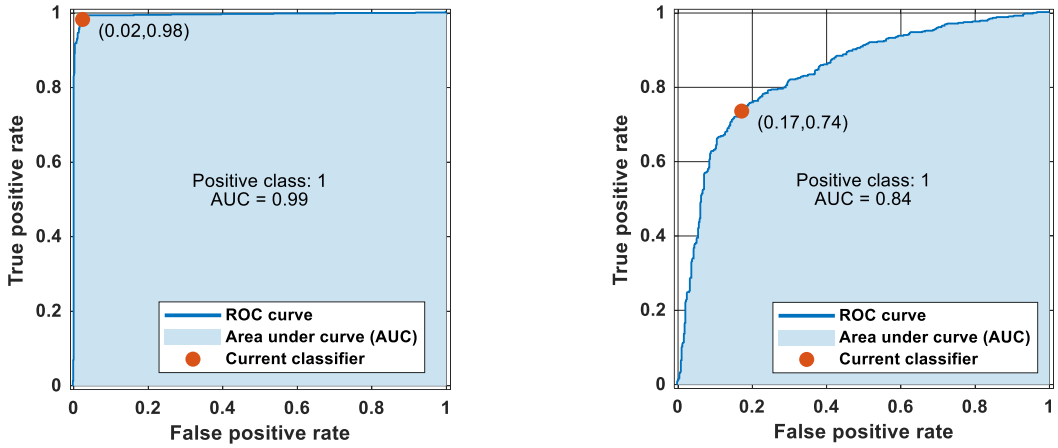
$$F - Measure = \frac{2TP}{2TP + FP + FN} \tag{12}$$

The confusion matrices obtained for the dataset are given in Figure 4 to comprehensively illustrate the calculated results.

a) b)

Figure 4. Confusion matrix a) ISH dataset b) Our collected dataset

In the confusion matrix given in Figure 4, the number 1 class refers to "Normal image" and the 2nd class refers to the "Spam image" images. In Figure 4.a, “Normal image” classification result for the ISH dataset is 98.39%, “Spam image” classification result is 98.71%. For the dataset we collected, the “Normal image” and “Spam image” classification results were calculated as 74.75% and 83.46%, respectively. ROC curves and AUC results obtained in the proposed method are shown in Figure 5.



a) b)

Figure 5. ROC curve of proposed method a) ISH dataset b) Our collected dataset

Figure 5 shows the ROC curves for the ISH dataset and the dataset we collected. The AUC value for the ISH dataset was 0.99, and the AUC value was obtained as 0.84 with the proposed method for our collected dataset. The proposed method was run for 1000 iterations for two datasets, and Accuracy, Precision, Recall, Geometric mean, and F Score values were calculated. The performance results obtained are listed in Table 3.

Table 3. Performance results are calculated by running 1000 iterations of the 10-fold-cross-validation method

Dataset	Statistic	Accuracy (%)	Precision (%)	Recall (%)	Geometric Mean (%)	FScore (%)
ISH dataset	Max	98.56	98.55	98.55	98.55	98.55
	Min	97.81	98.15	98.15	98.15	98.15
	Avg	98.21	98.52	98.51	98.51	98.51
	Std	0.11	0.04	0.04	0.04	0.04
Our collected dataset	Max	79.01	79.25	79.11	78.99	79.18
	Min	75.86	77.82	77.64	77.49	77.73
	Avg	77.48	78.97	78.84	78.73	78.91
	Std	0.46	0.23	0.23	0.23	0.23

As can be seen in Table 3, the highest accuracy was calculated as 98.56% for the ISH dataset. Maximum Accuracy, Precision, Recall, Geometric mean, and FScore values for our collected dataset were calculated as 79.01%, 79.25%, 79.11%, 78.99%, and 79.18%, respectively. The results of the proposed method were obtained by 10-fold cross-validation. The fold-wise accuracy results of the proposed method are presented in Figure 6.

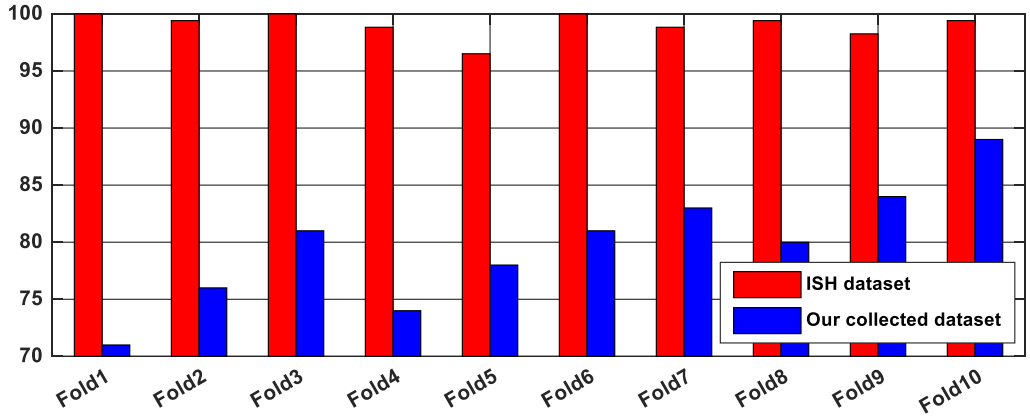


Figure 6. Fold-wise accuracy in percentage (%)

In the proposed method, the fold-wise accuracy results for the ISH dataset are over 95%. For our collected dataset, the lowest accuracy was calculated with Fold1 and the highest accuracy was calculated with Fold 10. In general, fold-wise results for Our collected dataset are calculated in the range of 70% to 90%.

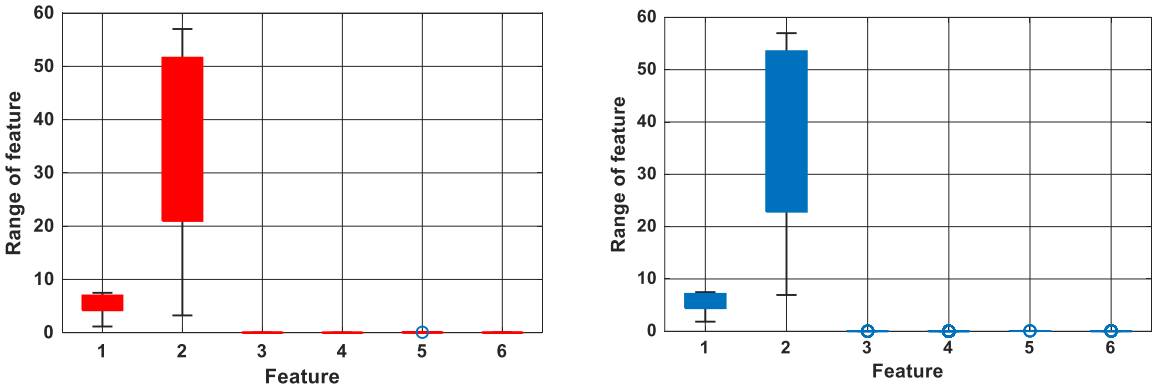
4. Discussion and Conclusions

In this study, a lightweight method was developed, and high performance was achieved on spam images. To test the proposed method, the dataset we collected was used in addition to the ISH dataset. The performance of the results obtained is compared with the studies in the literature in Table 4.

Table 4. Comparison of the proposed method and the literature for ISH dataset

Studies	Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Kumar et al. (Kumar vd., 2018), 2018	CNN	91.7	100	85.7	92.3
Salih et al. (Mahdi Salih & Nadeem Dhannoon, 2020), 2020	XYZ + CNN	98.4	-	-	-
Annadatha et al. (Annadatha & Stamp, 2018), 2016	PCA + SVM	97	-	-	-
Singh et al. (Singh vd., 2022), 2022	CNN	99.77	-	-	-
Belkhouche (Belkhouche, 2022), 2022	Byte histogram + CNN	94.85	95.23	94.75	94.83
Kihal et al. (Kihal & Hamza, 2023), 2023	VTA-CNN-RF	98.33	98.93	97.13	98.02
Ghizlane et al. (Ghizlane vd., 2022), 2022	CNN with CBAM	98.65	-	-	-
	CNN without CBAM	88.24	-	-	-
Our Method	**LBP + Statistic + SVM**	**98.56**	**98.55**	**98.55**	**98.55**

As can be seen in Table 4, the results of the proposed method are better than the literature. Kumar et al. (Kumar vd., 2018), classified spam images for the ISH dataset with the CNN algorithm. They calculated 91.7% accuracy. Salih et al. (Mahdi Salih & Nadeem Dhannoon, 2020), achieved 98.4% accuracy with XYZ image processing and CNN algorithm. For the ISH dataset, deep learning and machine learning-based methods and the proposed method were compared. The fact that the proposed method is both lightweight and high accuracy shows the advantages of the proposed method. Box chart graphs of the features selected in the proposed method are shown in Figure 7.



a) b)

Figure 7. Box chart denotation of the selected features a) ISH dataset b) Our collected dataset

In this study, we proposed a machine learning method to distinguish spam images from images in regular e-mails. The proposed method extracts image features and classifies them as spam or not. It was generated in the MATLAB program, and the results were computed. A total of 2754 photos were collected to test the procedure.

Moreover, these datasets demonstrate our proposal's overall success. The project outperformed existing image processing and machine learning approaches with an accuracy of 98.56%. These results and comparisons demonstrate that the proposed deep-feature extraction-based image spam detection method is capable of distinguishing between spam and non-spam e-mails.

References

Abuzaid, N. N., & Abuhammad, H. Z. (2022). Image SPAM Detection Using ML and DL Techniques. *International Journal of Advances in Soft Computing and its Applications*, *14*(1), 226-243. https://doi.org/10.15849/IJASCA.220328.15

Annadatha, A., & Stamp, M. (2018). Image spam analysis and detection. *Journal of Computer Virology and Hacking Techniques*, *14*(1), 39-52. https://doi.org/10.1007/s11416-016-0287-x

Belkhouche, Y. (2022). A language processing-free unified spam detection framework using byte histograms and deep learning. *2022 Fourth International Conference on Transdisciplinary AI (TransAI)*, 83-86. https://doi.org/10.1109/TransAI54797.2022.00021

Bhuiyan, H., Ashiquzzaman, A., Juthi, T. I., Biswas, S., & Ara, J. (2018). A Survey of Existing E-Mail Spam Filtering Methods Considering Machine Learning Techniques. *Global Journal of Computer Science and Technology: C Software and Data Engineering*, *1*(2). http://creativecommons.

Budanović, N. (2021). What's On the Other Side of Your Inbox – 20 SPAM Statistics for 2021. *DataProt*, 1.

Çayır, A., Yenidoğan, I., & Dağ, H. (2018). Feature Extraction Based on Deep Learning for Some Traditional Machine Learning Methods. *UBMK'18 3rd International Conference on Computer Science and Engineering*, 494-497.

Gao, Y., Yang, M., Zhao, X., Pardo, B., Wu, Y., Pappas, T. N., & Choudhary, A. (2008). Image spam hunter. *ICASSP 2008*, 1765-1768.

Ghizlane, H., Jamal, R., Mahraz, M. A., Ali, Y., & Hamid, T. (2022). Spam image detection based on convolutional block attention module. *2022 International Conference on Intelligent Systems and Computer Vision, ISCV 2022*, 0-3. https://doi.org/10.1109/ISCV54655.2022.9806065

Kihal, M., & Hamza, L. (2023). Robust multimedia spam filtering based on visual, textual, and audio deep features and random forest. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-023-15170-x

Kumar, A. D., R, V., & KP, S. (2018). DeepImageSpam: Deep Learning based Image Spam Detection. *arXiv*. http://arxiv.org/abs/1810.03977

Mahdi Salih, A., & Nadeem Dhannoon, B. (2020). Color Model Based Convolutional Neural Network for Image Spam Classification. *Al-Nahrain Journal of Science*, *23*(4), 44-48. https://doi.org/10.22401/anjs.23.4.08

Metlapalli, A. C., Muthusamy, T., & Battula, B. P. (2022). Classification of Image Spam Using Convolution Neural Network. *Traitement Du Signal*, *39*(1), 363-369. https://doi.org/10.18280/ts.390138

Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(7), 971-987. https://doi.org/10.1109/TPAMI.2002.1017623

PietikÃ¤inen, M. (2010). Local Binary Patterns. *Scholarpedia*, *5*(3), 9775. https://doi.org/10.4249/scholarpedia.9775

Rusland, N. F., Wahid, N., Kasim, S., & Hafit, H. (2017). Analysis of Naïve Bayes Algorithm for Email Spam Filtering across Multiple Datasets. *IOP Conference Series: Materials Science and Engineering*, *226*(1), 1-9. https://doi.org/10.1088/1757-899X/226/1/012091

Singh, A. B., Singh, K. M., Chanu, Y. J., Thongam, K., & Singh, K. J. (2022). An Improved Image Spam Classification Model Based on Deep Learning Techniques. *Security and Communication Networks*, *2022*, 1-11. https://doi.org/10.1155/2022/8905424

Trivedi, S. K. (2016). A study of machine learning classifiers for spam detection. *2016 4th International Symposium on Computational and Business Intelligence, ISCBI 2016*, 176-180. https://doi.org/10.1109/ISCBI.2016.7743279

Yasar, H., & Ceylan, M. (2021). A new deep learning pipeline to detect Covid-19 on chest X-ray images using local binary pattern, dual tree complex wavelet transform and convolutional neural networks. *Applied Intelligence*, *51*(5), 2740-2763. https://doi.org/10.1007/s10489-020-02019-1

Zhang, Z., Damiani, E., Hamadi, H. A., Yeun, C. Y., & Taher, F. (2022). Explainable Artificial Intelligence to Detect Image Spam Using Convolutional Neural Network. *2022 International Conference on Cyber Resilience (ICCR)*, 1-5. https://doi.org/10.1109/ICCR56254.2022.9995839